Research Article | Vol. 8, No. 11 | 1 Nov 2017 | BIOMEDICAL OPTICS EXPRESS 4838
Biomedical Optics EXPRESS

Novel health monitoring method using an RGB camera

M. A. HASSAN,[1,2] A. S. MALIK,[1,*] D. FOFI,[2] N. SAAD,[1] AND F. MERIAUDEAU[1]

[1]*Centre for Intelligent Signal and Imaging Research (CISIR), Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak, Malaysia*
[2]*Le2i UMR 6306, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté 12, rue de la Fonderie 71200 Le Creusot, France*
[*]*aamir_saeed@petronas.com.my*

Abstract: In this paper we present a novel health monitoring method by estimating the heart rate and respiratory rate using an RGB camera. The heart rate and the respiratory rate are estimated from the photoplethysmography (PPG) and the respiratory motion. The method mainly operates by using the green spectrum of the RGB camera to generate a multivariate PPG signal to perform multivariate de-noising on the video signal to extract the resultant PPG signal. A periodicity based voting scheme (PVS) was used to measure the heart rate and respiratory rate from the estimated PPG signal. We evaluated our proposed method with a state of the art heart rate measuring method for two scenarios using the MAHNOB-HCI database and a self collected naturalistic environment database. The methods were furthermore evaluated for various scenarios at naturalistic environments such as a motion variance session and a skin tone variance session. Our proposed method operated robustly during the experiments and outperformed the state of the art heart rate measuring methods by compensating the effects of the naturalistic environment.

OCIS codes: (330.0330) Vision, color, and visual optics; (330.7326) Visual optics, modeling.

References and links

1. M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," Opt. Express **18**(10), 10762–10774 (2010).
2. M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," IEEE Trans. Biomed. Eng. **58**(1), 7–11 (2011).
3. M. Lewandowska, J. Rumiski, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcam–a non-contact method for evaluating cardiac activity," in: Computer Science and Information Systems, 2011 Federated Conference on, (IEEE, 2011), pp. 405–410.
4. S. Kwon, H. Kim, and K. S. Park, "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone," in: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE, 2012), pp. 2174–2177.
5. T. Pursche, J. Krajewski, and R. Moeller, "Video-based heart rate measurement from human faces," in: 2012 IEEE International Conference On Consumer Electronics (IEEE, 2012), pp. 544–545.
6. H. Monkaresi, R. Calvo, and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam," IEEE J. Biomed. Health Informatics **18**(4), 1153–1160 (2014).
7. Y.-P. Yu, P. Raveendran, C.-L. Lim, "Heart rate estimation from facial images using filter bank," in: Communications, Control and Signal Processing, 2014 6th International Symposium on (IEEE, 2014), pp. 69–72.
8. X. Li, J. Chen, G. Zhao, M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2014), pp. 4264–4271.
9. A. Lam, Y. Kuno, "Robust heart rate measurement from video using select random patches," in: Proceedings of the IEEE International Conference on Computer Vision, (IEEE, 2015), pp. 3640–3648.
10. M. Kumar, A. Veeraraghavan, A. Sabharwal, and Distanceppg, "Robust non-contact vital signs monitoring using a camera," Biomed. Opt. Express **6**(5), 1565–1588 (2015).
11. L. Feng, L.-M. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," IEEE Trans. Circuits and Systems for Video Technology **25**(5), 879–891 (2015).
12. S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2016), pp. 2396–2404.
13. M. A. Hassan, A. S. Malik, D. Fofi, N. M. Saad, Y. S. Ali, and F. Meriaudeau, "Video-based heartbeat rate measuring method using ballistocardiography," IEEE Sensors J. **17**(14), 4544–4557 (2017).

#292965 https://doi.org/10.1364/BOE.8.004838
Journal © 2017 Received 17 Apr 2017; revised 16 Jul 2017; accepted 17 Jul 2017; published 4 Oct 2017

Biomedical Optics EXPRESS

14. G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," IEEE Trans. Biomed. Eng. **60**(10), 2878–2886 (2013).
15. A. R. Guazzi, M. Villarroel, J. Jorge, J. Daly, M. C. Frise, P. A. Robbins, and L. Tarassenko, "Non-contact measurement of oxygen saturation with an rgb camera," Biomed. Opt. Express **6** (9), 3320–3338 (2015).
16. C. E. Matthews, M. Hagströmer, D. M. Pober, and H. R. Bowles, "Best practices for using physical activity monitors in population-based research," Med. Sci. Sports Exerc. **44**(1 Suppl 1), S68 (2012).
17. W. Wang, S. Stuijk, and G. de Haan, "A novel algorithm for remote photoplethysmography: Spatial subspace rotation," IEEE Trans. Biomed. Eng. **63**(9), 1974–1984 (2016).
18. W. Wang, B. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote-ppg," IEEE Trans. Biomed. Eng.**64**(7), 1479–1491 (2017).
19. Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video Processing and Communications*, Vol. 5 (Prentice Hall, 2002).
20. A. Krishnaswamy and G. V. Baranoski, "A study on skin optics," Natural Phenomena Simulation Group, School of Computer Science, University of Waterloo, Canada, Technical Report. **1**, 1–17 (2004).
21. T. Lister, P. A. Wright, P. H. Chappell, "Optical properties of human skin," J. Biomed. Opt. **17** (9), 0909011 (2012).
22. G. R. Cooper, C. D. McGillem, *Probabilistic Methods of Signal and System Analysis* (Oxford University Press, 1986).
23. S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way* (Academic Press, 2008).
24. A. Antoniadis, "Wavelets in statistics: a review," Journal of the Italian Statistical Society **6**(2), 97–130 (1997).
25. W. Wang, S. Stuijk, and G. De Haan, "Exploiting spatial redundancy of image sensor for motion robust rppg," IEEE Trans. Biomed. Eng. **62**(2), 415–425 (2015).
26. M. Aminghafari, N. Cheze, and J.-M. Poggi, "Multivariate denoising using wavelets and principal component analysis," Computational Statistics & Data Analysis **50**, 2381–2398 (2006).
27. S. Mallat, *A Wavelet Tour of Signal Processing*, (Academic Press, 1999).
28. S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," IEEE Trans. Information Theory **38**(2), 617–643 (1992).
29. P. J. Rousseeu, and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," Technometrics **41**(3), 212–223 (1999).
30. J. Shi and C. Tomasi, "Good features to track," in: Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on (IEEE, 1994), pp. 593–600.
31. L. Scalise, N. Bernacchia, I. Ercoli, P. Marchionni, "Heart rate measurement in neonatal patients using a webcamera," in: Medical Measurements and Applications Proceedings, 2012 IEEE International Symposium on (IEEE, 2012), pp. 1–4.
32. M. A. Haque, R. Irani, K. Nasrollahi, and T. B. Moeslund, "Heartbeat rate measurement from facial video," IEEE Intelligent Systems **31**(3), 40–48 (2016).
33. M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," IEEE Transactions on Affective Computing **3**(1), 42–55 (2012).

1. Introduction

Remote health monitoring is a rapidly growing research field. Remote heart rate measurement using RGB camera have contributed vastly towards the growth of remote health monitoring. Change in the heartbeat rate would directly relate to the physiological/ pathological state of a person such as behavior, mood and stress. The heart rate measurement combined with modalities like EEG and MRI can result in improved accuracy of the diagnosis or analysis of brain conditions. Poh et al. [1] in 2010 reported the first facial video-based remote heart rate measuring method by using blind source separation (BSS) approach to estimate the photoplethysmography (PPG) signal from the red, green and blue spectrum of the RGB camera. Later researchers [2–7, 13, 32] presented a number of studies to improve the PPG signal quality by proposing temporal filtering methods, ROI selection and tracking methods, as well as heart rate estimation methods.

The BSS approach showed a limitation in component selection and motion artifact removal. Therefore, researchers proposed to estimate the PPG signal from the green spectrum of the RGB camera [8, 9] or from the monochrome camera with green filter [10]. The PPG signal estimation was also proposed by using a red and green spectrum of the RGB camera [11]. Alternatively, researchers also proposed to apply ill-pose inverse problem approach to estimate the PPG signal by using chrominance colors [12] by performing singular value decomposition and extracting the PPG signal from the largest singular value.

The majority of the researchers have measured the heart rate from the PPG signal by estimating the first harmonic/fundamental harmonic of the power spectral density of the PPG signal. Remote

Biomedical Optics EXPRESS

heart rate measurement at realistic conditions such as; uncontrolled illumination and motion settings have remained as a great challenge [12]. Heart rate is one of the main vital signs to access the physiological and pathological state of an individual. However, heart rate alone may not be a sufficient parameter to estimate the physiological and pathological state of an individual. The respiratory and autonomic nervous system parameters [15] derived from the heart rate variability could improve the assessment quality.

Furthermore, with the advancement in remote health monitoring; the proficiency to perform short-term heart rate measurement/health monitoring [16] would be the next step towards improving the efficiency of RGB camera based heart rate measuring methods. Currently, most methods with the exception of [10, 17], operate by estimating an average heart rate over a period of 30 seconds. Reducing the data size would hinder the reliability of the heart rate estimation.

Therefore, we present a novel heart rate and respiratory rate measurement method to operate on short-term period with smaller data size by estimating multivariate PPG signals (see Section 3) to overcome the issues of motion artifacts. The proposed multivariate PPG signal estimation method was developed after understanding and modeling the relationship of the skin reflectance and video signal generation to estimate the PPG signal (see Section 2). Upon estimating the heart rate and the respiratory rate, an experiment was conducted to benchmark and validate the proposed method (see Section 4). The results and discussion is presented in Section 5 and the study is concluded in Section 6.

2. Photoplethysmography model

Facial video based photoplethysmography (PPG) signal estimation, mainly relies on the quasiperiodic variation of the gradient intensity values of the red, green and blue spectra of the visible light spectrum. Previous studies, [1, 9–11, 17] have treated the quasiperiodic estimation of the gradient intensity values as a blind source separation problem or as an ill-posed inverse problem. Previously, researchers [10, 11, 14, 18] have modeled the principle of PPG extraction as a skin reflection model. Existing models have contributed to understanding the surface and subsurface reflectance of the light from the skin surface.

However, the relationship of the skin reflectance model to measuring video signal seems unclear. The main idea of PPG is formulated by measuring the illumination (i.e. gradient intensity) changes of the skin subsurface of the real world and projected to a image plane (i.e. video frame). After which estimating the time-varying temporal signal by projecting a series of 2-D image planes to a 1-D signal. Therefore, it is reasonable to say that PPG measurement is performed by video signal processing.

Therefore, understanding the PPG measurement from video signal processing by using an illumination model would be helpful to perform computer vision and machine learning operations. An illumination model describes the interaction of incident light with an object and its influence on the reflected light distribution (see Fig. 1). Here a spectral illumination model is used; since the aim is to model the relationship between the surface illumination variance in pixel intensity change of the video signal.

The interaction of light with the object surface can be described by three main entities of energy; incident flux, incident irradiance and reflected radiance [19]. The incident flux is known as the rate of which the energy is emitted from the light source. The incident irradiance is known as the incident flux per unit surface area of the object surface. The reflected radiance is the light energy that is reflected from the object surface.

Therefore the distribution of the reflected radiance C (i.e. the distribution of the illumination) can be modeled as a dependent (see Eq. (1)) of the incident irradiance E and the surface reflectance function r.

$$C(L, V, N, X, t, \lambda) = r(L, V, N, X, t, \lambda).E(L, N, X, t, \lambda) \tag{1}$$

Biomedical Optics EXPRESS

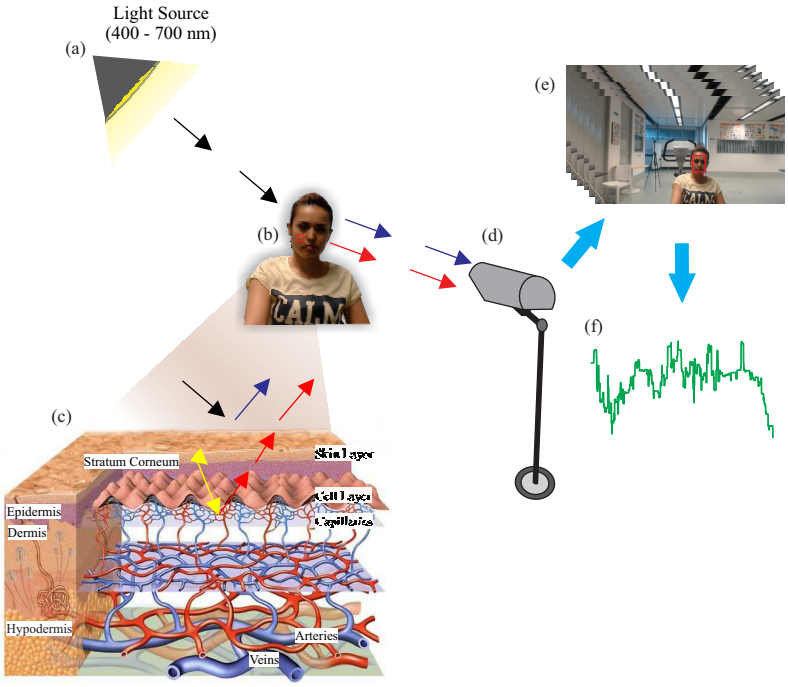


Fig. 1. Illustration of the 3-D to 1-D video signal processing of the Photoplethysmography Model. The black arrows highlight the incident light, the blue arrows highlight the surface reflectance and the red arrows highlight the subsurface reflectance. The green signal is the 1-D video signal.

where L is the direction of the illumination from the illuminating source, X is the real world location of the object surface. N is the surface normal vector at the location X. V is the viewing direction that connects the real world location X to the focal point of the camera. λ is the wavelength of the light that is emitted by the illuminating source.

The direction of the illumination L, viewing direction V and the surface normal vector N are functions of the real world location X and time t. The reflectance functions r is dependent on the wavelength λ of the incident light, the surface geometry and the material property of the surface.

Considering our aim to model the relationship of PPG measurement from facial skin; the facial skin can be taken as real world object location X that is captured by the video camera (see Fig. 2). Therefore, Eq. (1) can be simplified based on the assumptions that illumination source L and the viewing direction (i.e. camera position) are fixed (see Eq. (2)).

$$C(N, X, t, \lambda) = r(N, X, t, \lambda).E(N, X, t, \lambda) \tag{2}$$

The incident irradiance E would remain in Eq. (2), since the PPG measurement model is time varying and it is important to account for the motion of the object (i.e. rigid and non-rigid motion). Since we considered the skin as the world location; the reflectance functions r, of the object surface, can be derived as the function of the surface and subsurface reflectance [10, 14, 18].

$$r(t, \lambda) = r_s(t, \lambda) + r_d(t, \lambda) \tag{3}$$

Here r_s is the surface reflectance from the uppermost layer of the skin (i.e. stratum corneum). Surface reflectance can be accounted to cause the least variation to the distribution of the reflected radiance since the stratum corneum account to a negligible amount of absorption for the visible

Biomedical Optics EXPRESS

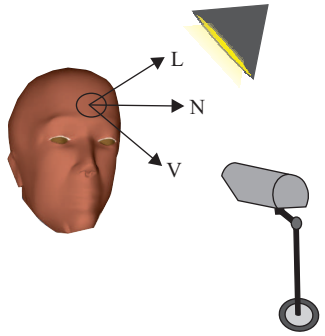


Fig. 2. Description of PPG measurement from facial skin by considering fixed L and V.

light spectrum [20]. The subsurface reflectance r_d is mainly contributed by the scattering at the epidermis, dermis, and hypodermis [20, 21]. The melanin is the main absorbent chromophore of the epidermis layer. The hemoglobin is the main absorbent chromophore of the dermis layer [20].

The hypodermis layer does not account for any significant absorption of the visible light. Therefore the visible light that reaches the hypodermis layer usually reflects back to the upper layers. However, melanin results to a constant absorption rate, where hemoglobin produces a variable absorption rate due to the blood volume pressure change (i.e. blood flow rate) in the capillaries. Therefore, the subsurface reflectance can be considered as the function of the blood volume pressure change $p(t)$ to the strength α of the modulation of the light backscattering from the subsurface (see Eq. (4)).

$$r_d(t, \lambda) = \alpha.p(t) \tag{4}$$

where α depends on the wavelength of the light, type of hemoglobin, concentration of melanin and also depends on the density of the capillary bed beneath the skin and orientation of the light source with respect to the surface of the skin. Therefore, the reflected radiance distribution C can be re-written (see Eq. (5)) as a function of the surface and subsurface reflectance that contain the information of the blood volume pressure change $p(t)$. Hence the video signal ψ of an observed skin surface can be modeled as the function of the reflected radiance distribution C and the spectral response function of the camera $a_c(\lambda)$ (see Eq. (6)).

$$C(N, X, t, \lambda) = (r_s(t, \lambda) + \alpha.p(t)).E(N, X, t, \lambda) \tag{5}$$

$$\psi(x, t) = \int_{\lambda=400}^{700} C(N, X, t, \lambda).(a_c(\lambda) + q(t))d\lambda \tag{6}$$

where $\psi(x, t)$ is the 2-D video signal that is dependent on time, x is the 2-D representation of the real world object surface. Here $x = P[X]$, $P[.]$ is the projection operator that projects the real world to the 2-D plane (i.e. image frame) and $q(t)$ quantization noise. The integration limits are set from $400nm$ to $700nm$ since the model is developed to understand the PPG estimation from RGB camera.

The 1-D video signal $\psi(t)$ is generated by estimated the spatial mean of the 2-D plane (see Eq. (7)). The spatial mean estimation process is the most commonly used process to transform the 2-D signal to 1-D since the spatial mean operation minimizes the quantization noise.

$$\psi(t) = \frac{1}{K \times M} \sum_{i=k}^{K} \sum_{j=m}^{M} x_{i,j}(t) \tag{7}$$

Biomedical Optics EXPRESS

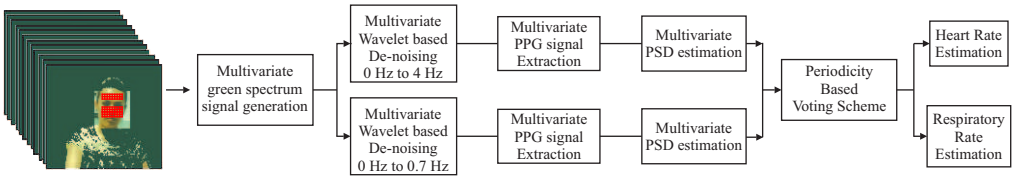


Fig. 3. Overall description of the framework to measure heartbeat rate using the Photoplethysmography model.

where K and M are the size of the horizontal and vertical axis of the ROI (region of interest) $x_{i,j}$.

3. Proposed framework

The heart rate and respiratory rate measurements are performed by exploiting the spatial-temporal periodicity of the PPG signal in the video signal $\psi(t)$. The proposed framework operates by detecting multiple ROI's from the spatial information of the face (see Fig. 3) to generate spatially separated H dimensional multivariate signals. The proposed framework only uses the green spectrum of the RGB camera since the green spectrum consists of the best tradeoff between a high absorption coefficient with hemoglobin [21] and penetration of the photons beneath the skin. The PPG signal is extracted by performing multivariate de-noising by using discrete wavelet decomposition.

The discrete wavelet decomposition is used since wavelets are able to segment the signals to different frequency constituents, and study each constituent with a resolution corresponding to its scale. Therefore the PPG signal can be extracted by decomposing and de-noising the multivariate signals of f_s sampling frequency to J number decompositions. The proposed framework estimate the heart rate and the respiratory rate by using a periodicity based voting scheme. The heart rate frequency and the respiratory rate frequency are estimated by selecting the frequency with the highest vote. The main idea of using the voting scheme is due to the fact that the PPG signal $P(t)$ that is created from the blood volume pressure change $p(t)$ (see Eq. (6)) is quasi periodic and the frequency of noise due to motion artifacts is random. Therefore, the heart rate frequency and the respiratory rate frequency in the PPG signal would repeat a greater number of times.

3.1. Multivariate signal generation

The 1-D video signal $\psi(t)$ is time dependent and is a 1-D representation of the function of the reflected radiance distribution C and the spectral response function of the camera $a_c(\lambda)$. The digital camera represents the visible light information as red, green, and blue; where information related to $\int_{\lambda=425}^{525} \lambda_r$ as blue spectrum, $\int_{\lambda=520}^{620} \lambda_g$ as green spectrum and the $\int_{\lambda=560}^{700} \lambda_r$ as the red spectrum. The literature related to skin optics shows that green spectrum has better tradeoff between a high absorption coefficient with hemoglobin in the dermis [21] and penetration of the photons beneath the skin. Therefore, would provide a stronger modulation α to the backscattering of the subsurface reflectance.

Before estimating the PPG signal $P(t)$ from the $\psi(t)$, it would be useful to understand the profile of the noise present in the $\psi(t)$ signal. An autocorrelation operation [22] was performed on the video signal $\psi(t)$ and the lag time plot (see Fig. 4) showed that $\psi(t)$ has a Gaussian noise profile. Therefore, the PPG signal estimation can be formulated as a de-noising problem [23, 24] (see Eq. (8)). Where $\varepsilon(t)$ represents the Gaussian noise of the video signal.

$$\psi(t) = P(t) + \varepsilon(t) \tag{8}$$

Biomedical Optics EXPRESS

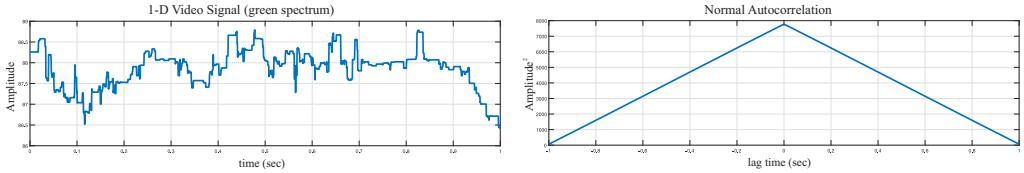


Fig. 4. Illustration of the noise profile of $\psi(t)$ by using normal autocorrelation.

The de-noising problem formulation may seem trivial and could be solved by using non-linear filtering methods. However, noise parameter $\varepsilon(t)$ is time dependent and would vary with the change of time. As discussed in the PPG model (see Section 2) the PPG signal is quasi periodic in nature. Previous work [9, 10, 25] showed that the skin can be spatially separated into multiple regions and can be treated as independent regions. Therefore, the PPG signal $P(t)$ can be extracted by exploiting the spatial-temporal information of the video signal $\psi(t)$. Hence, the Eq. (7) can be transformed to a multivariate signal by spatially separating the facial skin region to multiple regions (see Eq. (9)).

$$\psi_h(t) = \frac{1}{K \times M} \sum_{h=1}^{H} \sum_{i=k}^{K} \sum_{j=m}^{M} x_{h,i,j}(t) \tag{9}$$

where H is the number of the spatially separated facial regions. The de-noising problem also can be transformed to a multivariate de-noising problem (see Eq. (10)).

$$\psi_h(t) = P_h(t) + \varepsilon_h(t), 1 \leqslant h \leqslant H \tag{10}$$

where $\psi_h(t)$, $P_h(t)$ and $\varepsilon_h(t)$ are H-dimensional multivariate signals. Therefore, a wavelet based multivariate de-noising operation [26] is performed to extract the PPG signal $P(t)$. The wavelet based multivariate de-noising approach operates mainly by decomposing $\psi_h(t)$ to J number of levels and removing the noise by a adaptive thresholding operation.

3.2. Wavelet based decomposition

During the de-noising phase, the multivariate signal were considered in discrete time since the acquired multivariate signal $\psi_h(t)$ is processed as samples [27, 28]. Therefore, the time variable t is replaced by n, where n is the number of samples. Discrete wavelet transform (DWT) operates in the discrete time space with a finite number of samples. DWT presents the application of multiresolution analysis. Multiresolution analysis of a signal is essential to derive the information present. The scaling function $\varphi(n)$, assists the multiresolution analysis by spanning the data/samples across a span of subspaces where $\varphi(n) \in L^2$.

However, spanning the data across subspaces is not sufficient, and it is required to span the difference between the spaces by using different scales of the scaling function. The different scales of the scaling function are the wavelets $\xi(n)$. By defining the wavelets function the wavelet vector space in the subspace can be derived as, $V_1 = V_0 \oplus W_0$; where V_0 is the initial space and W_0 is the mother wavelet. The DWT (i.e. forward) can be derived for the video signal $\psi(n)$, containing n number of samples as shown in Eq. (11) and Eq. (12).

$$W_\varphi(j_0, k) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \psi(n) \varphi_{j_0,k}(n) \tag{11}$$

Biomedical Optics EXPRESS

$$W_{\xi}(j,k) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \psi(n) \xi_{j,k}(n) \tag{12}$$

where N is the total number of samples and k is the time translation, and j is the scaling index. Furthermore, upon DWT a signal can be analyzed using the wavelets and the scaling function, where the wavelet and scaling functions can be derived as shown in Eq. (13) and Eq. (14).

$$\xi(n) = \sum_{k} h_{\xi}(k) \sqrt{2} \varphi(2n - k) \tag{13}$$

$$\varphi(n) = \sum_{k} h_{\varphi}(k) \sqrt{2} \varphi(2n - k) \tag{14}$$

where h_{ξ} is the wavelet function coefficient and h_{φ} is the scaling function coefficient. Here we used the symlet wavelet with two vanishing points as the wavelet function since the symlet has a pulse like shape similar to the PPG pulse. The wavelet transform decomposition of $\psi(n)$, can be decomposed into two forms which are the details and the approximates. Here $\xi(n)$ represents the details of the signal and $\varphi(n)$ represents the approximates of the signal. Considering the first decomposition, the details will represent the high-frequency content (i.e. variables that vary fast) and the approximates represent the low-frequency content (i.e. variables that vary slowly) of the signal.

For the second decomposition the approximates of the first decomposition was further decomposed into their details and approximates. The process was continued to J levels until the wavelet scale reached the heart rate frequency range 0.7 Hz to 4 Hz (i.e. 42 bpm to 240 bpm). Hence four levels of decomposition ($J = 4$) are carried out for a signal that is recorded at a sampling frequency f_s of 30 Hz (i.e. frame processing rate 30 fps). Similarly, the video signals were decomposed to six levels ($J = 6$) to capture the respiratory rate frequencies information between 0.2 to 0.7 Hz.

3.3. De-noising

The de-noising of the multivariate signal is performed by removing the noise elements ε by thresholding the noisy signals. The noisy elements are calculated by estimating a noise covariance matrix Σ_{ε}. The noise covariance matrix Σ_{ε} is estimated by computing the minimum covariant determinant [29] of the first details $W_{\xi}(1,n)$ of the multivariate signals. The details $W_{\xi}(1,n)$ of the first decomposition is used since $W_{\xi}(1,n)$ represent the high frequency content (i.e. variables that vary fast) of the video signal ψ. The noise covariance matrix Σ_{ε} decomposed into orthogonal matrix $\vee$ and diagonal matrix $\wedge$ (see Eq. (15)).

$$\Sigma_{\varepsilon} = \vee \wedge \vee^{T} \tag{15}$$

where $\wedge = diag(\mu_h, 1 \leqslant h \leqslant H)$. The de-noising is performed by soft thresholding γ_i, the H univariate signals of the orthogonally transformed details $W_{\xi}(j,n)\vee$ of each decomposition, where $W_{\xi}(j,n)\vee, 1 \leqslant j \leqslant J$. The soft thresholding is performed by using the diagonal elements of $\wedge$ (see Eq. (16)). The useful information in the J^{th} approximate $W_{\varphi}(J,k)$ is extracted by performing PCA and extracting the H_{J+1} principal components. Here the H_{J+1} principal components were selected to exclude redundant features and extract the most significant features. Finally the de-noised multivariate PPG signal $P_h(n)$ is extracted by reconstructing de-noised $\hat{\psi}_h(n)$, from the simplified details and approximates by using invert wavelet transforms.

$$\gamma_h = \sqrt{2\mu_h log(n)} \tag{16}$$

Biomedical Optics EXPRESS

Algorithm 1 Wavelet based Multivariate De-noising

procedure MULTIVARIATE VIDEO SIGNAL($\psi_h(n)$)
  for j = 1:J **do** ▷ Levels of decompositions
    $[\xi(n,j), \varphi(n,j)] = DWT(\psi_h(n))$ ▷ Signal decomposition using Discrete Wavelet Transform
  end for
  $\Sigma_\varepsilon = MCD(\xi(n,1))$ ▷ Noise covariance matrix estimation using Minimum Covariant Determinant
  $\Sigma_\varepsilon = \vee \wedge \vee^T$ ▷ Computing the orthogonal matrix
  for j = 1:J **do**
    $\xi_j \vee = \xi_j \times \vee$ ▷ Wavelet function Basis change
    for h = 1:H **do**
      $\hat{\xi}_{j,h} = ST(\xi_{j,h} \vee, \gamma_h)$ ▷ Univariate Soft Thresholding
    end for
  end for
  $\hat{\varphi} = PCA(\varphi)$ ▷ Scaling function Basis change
  $\hat{\psi}_h(n) = Reconstruct(\hat{\xi}, \hat{\varphi})$ ▷ Reconstructing de-noised matrix
  $p_h(n) = \hat{\psi}_h(n)$
  return $p_h(n)$ ▷ Multivariate PPG signal
end procedure

3.4. Heart rate estimation

The de-noised PPG signal $P_h(n)$ is multivariate and manually selecting the best possible PPG signal from the H number of signals would be inefficient. Previously the heart rate is measured by estimating the first harmonic frequency of the PPG signal. However, all de-noising methods do not extract the ideal PPG signal at all instances. Often the de-noised PPG signal is affected by motion artifacts and may contain a frequency of the motion artifact with a higher power. However, the frequency of the heart beat does not disappear; rather the frequency of the heart beat is shifted to a frequency with lower power in the power spectral distribution (PSD).

Therefore, we estimated the heart rate by exploiting the periodicity of the PPG signal by a frequency based voting scheme. The PSD of the multivariate PPG signals was derived (see Eq. (17)), and we empirically selected the five frequencies with the highest peak (see Fig. 5) in the PSD for each H dimensional multivariate PPG (see Eq. (18)). The heart rate frequency f_{HR} (see Fig. 6) was selected as the frequency with the highest number of repetitions/votes.

For the case of multiple frequencies repeating the same number of time; we selected the frequency with the accumulative highest power of the H dimensional multivariate signals. The main idea of using a voting scheme based approach to estimate the heart rate frequency f_{HR}, was due to the fact that the heart rate frequency is periodic and the noise of motion artifacts are random. Therefore, the heart rate frequency would repeat higher number of time compare to the frequencies of the motion artifacts.

$$\phi(\omega) = \lim_{N\to\infty} E\{\frac{1}{N}|\sum_{n=0}^{N-1} p_h(n)e^{-i\omega t}|^2\} \tag{17}$$

$$H_f(\phi_{1:5,h}(\omega)) = \sum_{h=1}^{H}[\phi_{1,h}(\omega)\phi_{2,h}(\omega)\phi_{3,h}(\omega)\phi_{4,h}(\omega)\phi_{5,h}(\omega)] \tag{18}$$

where H_f is the matrix that capture the five frequencies with the highest power of the H number of PPG signal's.

Biomedical Optics EXPRESS

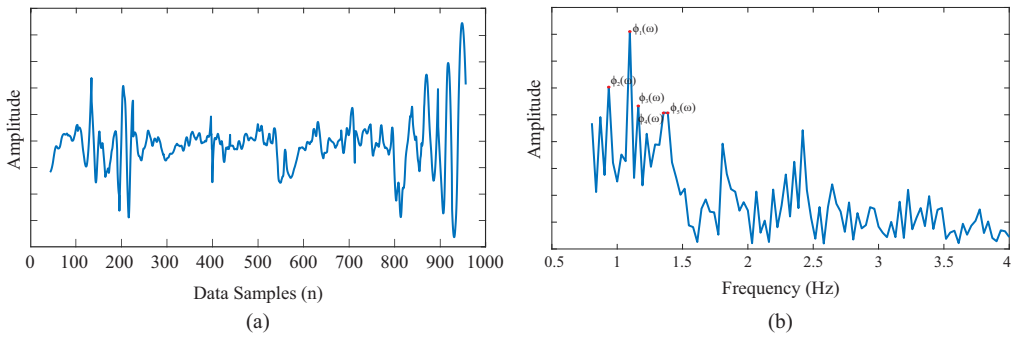


Fig. 5. Description of the first five harmonics selection from the (a) de-noised video signal from the (b) PSD.

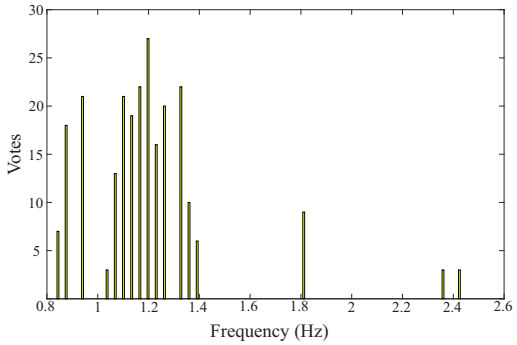


Fig. 6. Illustration of the heart rate estimation by voting scheme.

3.5. Respiratory rate estimation

The respiratory rate measurement is obtained based on the respiratory motion. The ROI used for the framework of the PPG signal estimation method was used to maximize the effect of respiratory motion that directly modulates the baseline of the PPG signal. Therefore, same de-noising framework was used to measure the respiratory rate. However, the operational frequencies for the respiratory rate were varied to 0.15 to 0.66 Hz corresponding to the respiratory rate of 9 to 40 respiration per minute (rpm).

Therefore, the video signal was decomposed to $J = 6$ levels for a $f_s = 30$ Hz video signal. The respiratory rate was also measured by using the periodicity based voting scheme (PVS) (see Section 3.4). However, we empirically selected only three frequencies with the highest power. The frequency with the highest number of votes was selected as the respiratory frequency (see Fig. 7(a) & 7(b)) and the respiratory rate was measured by multiplying the respiratory frequency by 60 to estimate the respirations per minute.

3.6. ROI detection and tracking

The ROI detection and tracking is an essential component to generate $\psi(n)$. As described in the PPG model (see Section 2), Eq. (2), considers that the facial skin region is captured and tracked along time. Therefore, we used Viola-Jones face detector to detect the face of the subject. The proposed framework to measure the PPG signal utilized multiple ROI's from two regions of the face. The first region involved the forehead of the face and the second region was extracted from the skin region below the eyes and above the upper lips of the face.

Biomedical Optics EXPRESS

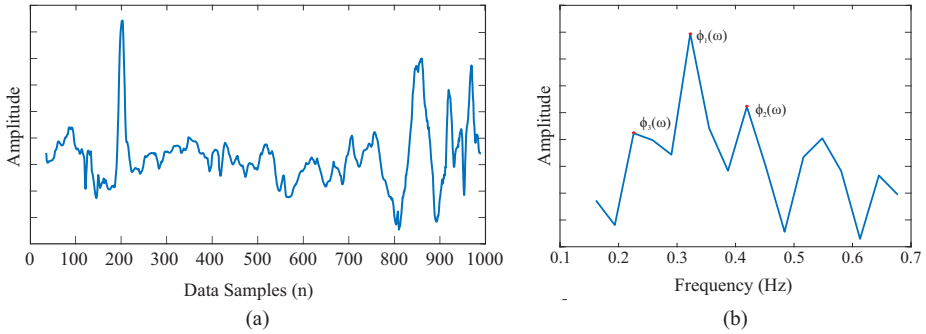


Fig. 7. Description of the first three harmonics selection from the (a) de-noised video signal from the (b) PSD.

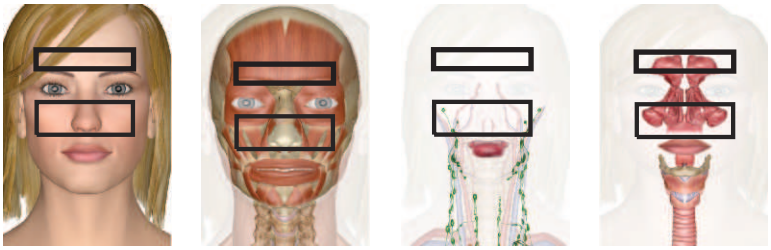


Fig. 8. Illustration of the ROI selection. (a) the ROI selection in prospect of outlier skin, (b) ROI selection in terms of tissue and muscle coverage, (c) ROI selection in terms of arteries, arterioles coverage and (d) ROI selection in terms of upper respiratory system coverage (i.e. sinus).

The two skin regions were mainly used since these two areas are least prone to facial hair. The facial hair may result as a motion artifact in the PPG signal. Furthermore, Fig. 8 shows that these facial regions contain denser capillary bed that would exhibit high pulsatility. The selected regions also cover the upper sinus of the respiratory system (see Fig. 8(d)). Therefore these regions are much liable to transmit a stronger respiratory motion.

The two regions were tracked temporally by using KLT feature tracking algorithm. The tracking strategy involved two-step processes; detect and track. Here the detection was performed by identifying $Z \simeq 50$ feature points for tracking based on the feature selection process of good features to track approach [30]. The KLT was used to track the good features. As the number of features reduces below 50, the feature detection step is re-initialized to detect the new batch of good features to track. The two step, detect and track approach was mainly used to overcome the issues related to ROI drift and ROI loss.

Taking the spatial mean of single or large ROI [1–3, 8, 11, 12, 31] often results on carrying motion artifacts caused by spatial illumination variance due to the non-symmetric nature of the facial skin surface. Therefore, our framework proposed to extract multiple ROI's from the two main regions that were extracted by using Viola-Jones face detector.

Therefore the forehead region is spatially separated to 16 ROI's and the mid facial region is spatially separated to 32 ROI's to capture the subtle variations of smaller facial regions (see Fig. 9(i)). Hence multiple ROI approach is able to minimize the effect caused by the spatial illumination variation of the face. The multiple ROI approach is essential to the framework as the multiple ROI time variant signals ($\psi_H(t)$) satisfies the formulation of the multivariate de-noising

Biomedical Optics EXPRESS

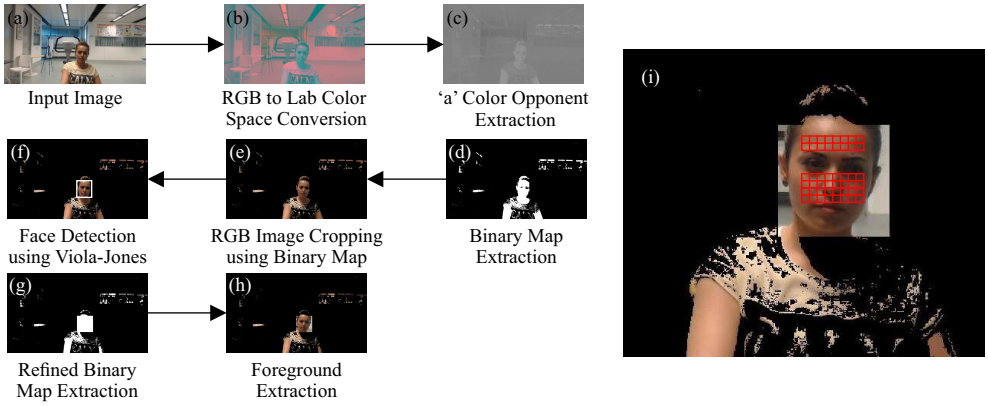


Fig. 9. Illustration of the ROI detection after performing foreground object detection. (a) Input Image/frame, (b) CIELab color space conversion, (c) 'a' color opponent extraction, (d) binary map extraction, (e) RGB image cropping using binary map, (f) Face detection using Viola Jones, (g) Refining binary map, (h) Foreground extraction and (i) multiple ROI detection.

problem (see Eq. (9)).

The face detection methods are not reliable for the naturalistic environment. Therefore, we utilized the foreground object detection based on segmenting skin color using CIELab color space as a preprocessing step [13]. The overall process of the foreground extraction and ROI detection is illustrated in Fig. 9)

4. Experiment setup

A series of experiments were conducted to evaluate the heart rate estimation of the proposed method along with the state of the art methods. The state of the art methods included the remote heart rate measuring methods proposed by Poh et al. [2], using BSS on RGB spectrum, de Haan et al. [14], using chrominance space derived from the RGB spectrum, Lam et al. [9], using green spectrum based BSS method, Feng et al. [11], using green and red differentiation operation, and Wang et al. [17], using spatial-subspace rotation of the RGB spectrum for PPG signal estimation. All the methods were implemented in MATLAB and run on a personal computer with a 3.4 GHz processor and 8 GB RAM. The experiments were carried out for two databases; Experiment 1, using publicly available MAHNOB-HCI database [33] and Experiment 2, using self-collected dataset at naturalistic environment.

Experiment 1: MAHNOB-HCI database is a multi-model database; we used the emotion elicitation experiment. During the experiment, we used 27 subjects of which 15 were female, and 12 were male of age 19 to 40 years old (mean of 26.06 and standard deviation of 4.39). We used the data, recorded with a AVI compression using a 24 bit RGB camera of 61 fps (frames processed per second) at a resolution of 780×580. The camera was positioned in front of the subject at a distance of 0.4 meters. The ground truth for the heart rate measurement was extracted from the ECG signal of channel 34 of the sensor attached to the upper left corner of the chest, under clavicle bone. We also used the MAHNOB-HCI database to validate the respiratory rate estimation. Here the ground truth for the respiratory rate measurement was extracted from the respiratory belt connected to channel 45.

Experiment 2: The naturalistic environment experiment was carried out to investigate the effect of real-world situations. Therefore, we tested our proposed method for three scenarios: neutral session, motion variance session and skin tone variance session. The experiments were carried out at workstation environment with no control to the illumination of the environment.

Biomedical Optics EXPRESS

The facility used for data collection was illuminated by fluorescent lamps that were fixed on the ceiling and the light from the surrounding environment. The non-controlled illumination produced an uneven illumination distribution across the face, causing shadows and specular reflections on the surface of the face, as illustrated in Fig. 11(b).

The experiments were conducted on 45 healthy subjects. The subjects were separated into three groups based on the skin tones (i.e. Fair, Brown and dark). The subject's age varied from 21 to 63 with a mean of 29.77 years and standard deviation of 7.88 years. We used the Logitech C920 HD Pro Webcam to record the video. The camera was a 24 bit RGB camera with Bayer mosaic filter that recorded uncompressed data at 30 fps. The camera operated at a resolution of 1080×1920 and was mounted at 0.8 meters from the subject. The ground truth heart rate was captured using a pulse oximeter. The WristOx2 model 3150 wrist-worn pulse oximeter by Nonin Medical.

5. Results and discussion

Following the experiment, the estimated heart rate M_{hr} was validated to the ground truth heart rate G_{hr}. The results were derived from absolute heart rate error H_e as shown in Eq. (19). The derived heart rate error was analyzed by using the mean of the heart rate error Me, standard deviation SDe, root mean square error $RMSE$ and Pearson correlation coefficient r. The results of our experiments (i.e. Mean Error, Root Mean Square Error and Pearson correlation coefficient) were compared with results of previous studies [8,9,12,32]. The similar error rates and/or Pearson correlation coefficient show that the state of the art was re-implemented correctly.

$$H_e = |M_{hr} - G_{hr}| \tag{19}$$

5.1. Experiment 1

The experiment 1 was conducted to evaluate the proposed method along with existing methods on the publicly available database. Here the experiments were carried out for two-time fragments: long-term average using 30s video and short-term average using 10s video. The results of the experiment are tabulated in Table. 1; the BSS approach using RGB spectrum [2] underperformed compared to the recently proposed heart rate measuring methods. The underperformance may have occurred due to the variation of the spectral response of the camera over time. Therefore, affecting the linear mixing coefficient of the R, G, B signals. The underperformance may have also occurred due to the video compression and the possible gamma correction of the videos in the database.

The video compression and the gamma correction often suppress the subtle color changes due to blood volume pressure change. Hence affecting the functionality of the multi-color BSS approach. The BSS approach using the green spectrum were able to improve the results by proposing a random selection of the facial patches by computing a confidence ratio to estimate the best matching patches. Lam et al. [9] showed substantial improvement over [2]. However, the random patch selection of [9] is computationally expensive and does not always select the patches with the same illumination spectrum rather selects the patch with the highest similarity.

The chrominance based method proposed by De Haan et al. [14] further improved the error between the ground truth and the RGB camera based approach. The meaningful representation of the skin pixel using chrominance features may have improved the PPG signal estimation. However, most of the methods failed to maintain the reliability during the short-term average heart rate estimation session. The videos used from the MAHNOB-HCI database were recorded at 61 fps. Therefore, capturing only 610 samples of data during the 10 seconds. The lack of data may have contributed to the underperformance of these data-driven methods.

The spatial subspace rotation approach using RGB spectrum proposed by Wang et al. [17] operated stably by reporting similar error rate during both long-term average and short-term

Biomedical Optics EXPRESS

average heart rate estimation sessions. However, the method underperformed for the MAHNOB-HCI database. The video compression and the possible gamma correction of the videos in the database could have also been the cause for the underperformance of the method. Our proposed method using multivariate de-noising approach was developed to operate for both long-term average and short-term average heart rate estimation.

The method mainly operated by exploiting the spatial periodicity of the multivariate raw PPG signals by performing wavelet based de-noising and using a PVS to estimate the heart rate. Therefore, the attributes of the de-noising algorithm and the PVS were both evaluated during the experiment. The results (see Table. 1) showed that the de-noising algorithm operated reliably for long-term average heart rate estimation. However, the error rate was further improved by including the PVS for both long-term average and short-term average heart rate estimation sessions.

The overall approach of not using a bandpass filter in our proposed method enabled the respiratory rate estimation at lower frequencies. The respiratory rate estimation was validated using Bland-Altman analysis and correlation estimation between the measured respiratory rate and the ground truth respiratory rate (see Fig. 10). The Bland-Altman plot showed a dense grouping for the difference between the measured respiratory rate and ground truth respiratory rate. The analysis derived a bias of 0.58 rpm for 95% limits of agreement between -4.34 rpm and 5.50 rpm. Similarly, the correlation plot showed a dense grouping with a Pearson correlation coefficient of 0.8 for a 95% confidence interval between 0.77 and 0.84. The R-squared indicated 0.65 of the respiratory rate measurements were closely fitted to the regression line.

Table 1. Performance validation of the methods for benchmarking experiment of 30s videos and 10s videos

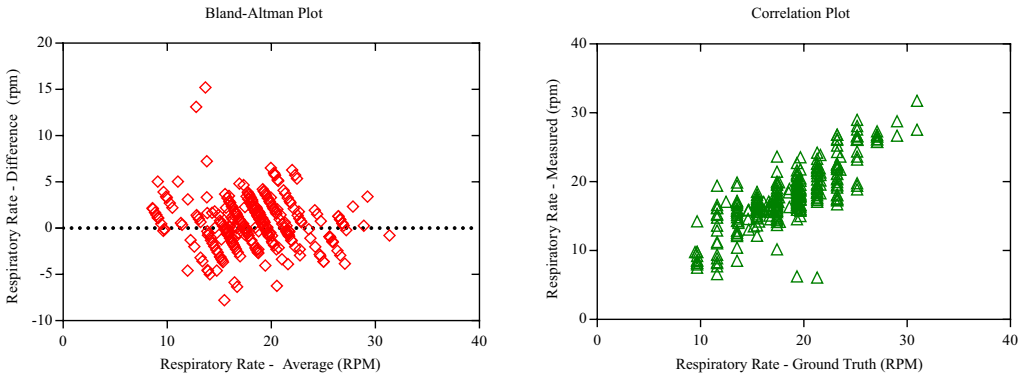
Methods	30s Video				10s Video			
	Me (bpm)	SDe (bpm)	RMSE (bpm)	r	Me (bpm)	SDe (bpm)	RMSE (bpm)	r
Poh et al. [2]	11.87	10.26	15.67	0.52	13.83	11.95	17.42	0.45
De Haan et al. [14]	7.41	5.86	8.48	0.82	10.71	8.45	13.64	0.61
Lam et al. [9]	8.73	4.94	10.02	0.81	8.11	6.03	10.07	0.68
Feng et al. [11]	8.05	7.04	10.36	0.62	8.54	9.58	12.76	0.55
Wang et al. [17]	7.48	4.76	9.11	0.81	7.37	7.48	10.48	0.69
Proposed without PVS	4.97	4.08	6.42	0.85	6.84	5.07	8.51	0.74
Proposed with PVS	**3.89**	**3.46**	**5.2**	**0.89**	**4.73**	**5.21**	**7.03**	**0.81**

5.2. Experiment 2

The experiment 2 was carried out to validate the performance of the heart rate estimation methods for naturalistic environment while the illumination of the scene is not controlled. Here the results for the neutral session, motion variance session and skin tone variance session for are tabulated in Table. 2. The results of the BSS approach [2] for the neutral session improved compared to the results of experiment 1 (see Table. 1). However, despite the improvement over the Pearson correlation coefficient, the method reported a high error rate. The challenge of spatial illumination variance from the naturalistic/uncontrolled environment (see Fig. 11) may have contributed to the higher error rate.

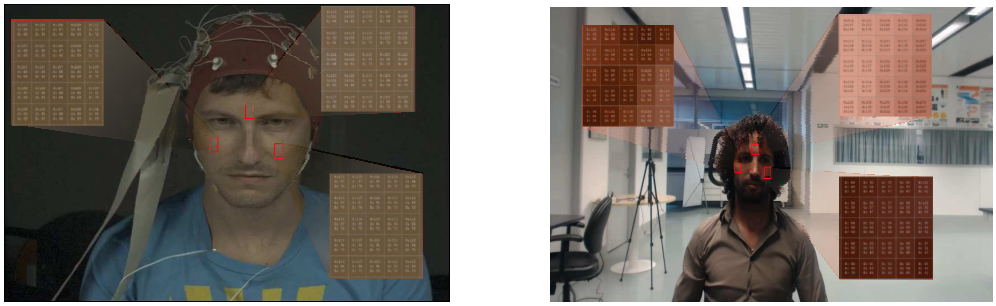
The illumination of the naturalistic environment is uncontrolled. Thus, the distribution of the illumination across the facial skin surface is uneven (see Fig. 11). Therefore, the approach of spatial averaging a large region may have contributed to underperformance in the naturalistic environment. Similarly, the error rate for the proposed method of [9] also could have increased due to the effect of spatial illumination variance. However, the proposed method of De Haan

Biomedical Optics EXPRESS



(a) (b)

Fig. 10. Performance validation of the respiratory rate estimation using Bland-Altman plot (a) and Correlation plot (b)



(a) (b)

Fig. 11. Illustration for the effect of spatial illumination variance at (a) controlled environment and (b) uncontrolled environment.

et al. [14] robustly at naturalistic environment. The transformation of the RGB space to the chrominance feature space may have contributed to overcome the effect of spatial illumination variance.

The spatial subspace rotation approach [17] improved the error rate compared to experiment 1 (see Table. 1) and operated robustly at the naturalistic environment. The formation of the spatial subspace for estimating the PPG sample may have contributed to overcoming the effect of the naturalistic environment. Our proposed operated robustly and was able to achieve similar results to experiment 1. The multi-ROI across the skin surface along with the multivariate de-noising approach was able to compensate the effects of the spatial illumination variance.

The results for the motion variance experiment showed that proposed method by Wang et al. [17] operated robustly compared to the state of the art methods. The approach of rotating and scaling the spatial subspace may have contributed to extracting the pulse of the PPG signal from the drift of the raw trace. However, our proposed method underperformed during the motion variance session. Here the subject exhibited rigid and non-rigid motion by changing posture and expressing emotions/talking. The inability of the multivariate de-noising method to rectify the temporal drift of the signal could have caused the rise in error rate.

The performance of the heart rate estimation methods were also evaluated for different skin tones. Most of the methods reported a lower error rate and a high Pearson correlation coefficient

Biomedical Optics EXPRESS

Table 2. Performance validation of the proposed method for naturalistic environment experiment

Session	Method	Me (bpm)	SDe (bpm)	RMSE (bpm)	r
Naturalistic Environment Experiment					
Neutral	Poh et al. [2]	11.04	8.19	13.69	0.64
	De Haan et al. [14]	7.32	5.02	8.84	0.84
	Lam et al. [9]	9.24	5.15	10.55	0.78
	Feng et al. [11]	9.11	5.68	10.70	0.76
	Wang et al. [17]	6.02	5.09	7.84	0.84
	Proposed	**3.52**	**3.54**	**4.97**	**0.89**
Motion variance	Poh et al. [2]	14.00	10.05	17.17	0.56
	De Haan et al. [14]	10.00	6.34	11.80	0.68
	Lam et al. [9]	10.51	8.08	13.20	0.66
	Feng et al. [11]	7.26	6.24	9.53	0.72
	Wang et al. [17]	**6.54**	**4.91**	**8.15**	**0.79**
	Proposed	4.62	5.75	7.28	0.74
Skin Tone Variance Experiment					
Fair Skin	Poh et al. [2]	10.16	6.67	12.03	0.7
	De Haan et al. [14]	6.55	4.15	7.55	0.85
	Lam et al. [9]	8.17	3.40	8.81	0.83
	Feng et al. [11]	8.99	4.78	10.10	0.82
	Wang et al. [17]	**3.79**	**2.86**	**4.69**	**0.88**
	Proposed	4.35	3.63	5.57	0.85
Brown Skin	Poh et al. [2]	12.07	6.67	13.68	0.58
	De Haan et al. [14]	7.32	5.79	8.44	0.84
	Lam et al. [9]	9.17	4.82	10.28	0.79
	Feng et al. [11]	7.92	6.49	10.10	0.75
	Wang et al. [17]	4.97	4.06	6.33	0.84
	Proposed	**4.86**	**3.43**	**5.86**	**0.88**
Dark Skin	Poh et al. [2]	10.89	10.94	15.18	0.61
	De Haan et al. [14]	**8.01**	**5.13**	**9.42**	**0.82**
	Lam et al. [9]	12.05	6.54	13.61	0.67
	Feng et al. [11]	10.42	5.75	11.81	0.65
	Wang et al. [17]	6.53	7.45	9.72	0.78
	Proposed	4.38	7.56	8.86	0.61

for fair skin tone, while [17] reporting the lowest error rate. Similarly, our proposed method, [14] and [17] reported lower error rate for brown skin tone. However, with the exception of [14], most of the methods underperformed for dark skin tones. The underperformance could have been caused by the low modulation strength α of the light backscattering from the subsurface of the skin. Since dark skin exhibit, a higher melanin concentration compared to fare and brown skin [20]. Therefore, affecting the PPG signal estimation process. However, the chrominance feature based approach [14] operated reliably for the skin tones variance session by reporting a high Pearson correlation coefficient across all three different skin tones.

We also evaluated the computational complexity of our PPG heart rate estimation method for experiment 2 using the profiler tool of MATLAB. The overall execution time measurement for the 30s video was 36.08s; resulting to an overall processing speed of 24.93 frames per second. Further analysis to the execution time showed that 34.73s were utilized for ROI detection and

Biomedical Optics EXPRESS

tracking component, and 0.39s for the multivariate de-noising component, and 0.06s for the periodicity based voting scheme component. Thus, the multivariate de-noising algorithm along with periodicity based voting scheme consuming only about 0.45s i.e. 1.25% of the overall execution time measurement.

6. Conclusion

In this paper, a health monitoring method was proposed by remotely measuring the heart rate and respiratory rate from the PPG signal. The method was developed as result of modeling and understanding the relationship of skin reflectance and video signal generation. The proposed method was formulated by estimating the PPG signal from the video signal by using a multivariate de-noising approach along with a periodicity based voting scheme to estimate the heart rate. The proposed method was validated with multiple experiments. The results showed that our method outperformed the state of the art method for long-term average using 30s video and short-term average using 10s heart rate estimation experiments and operated robustly during the naturalistic environment experiment. During our future work, we would address the issue of motion artifacts caused by the temporal drift of the de-noised PPG signal and extend our method to perform health screening by measuring multiple vital signs.

Funding

HiCoE grant for CISIR (Ref No. 0153CA-002), Ministry of Education (MOE), Malaysia.

Disclosures

The authors declare that there are no conflicts of interest related to this article.